Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 34 (2008) 119–127
Expert Systems with Applications
www.elsevier.com/locate/eswa

Using customer knowledge in designing electronic catalog

Chinho Lin [a,*], Chienwen Hong [a,b]

[a] Department of Industrial and Information Management, College of Management, National Cheng Kung University, No. 1, Ta-Hsueh Road, Tainan 701, Taiwan, ROC
[b] Department of Information Management, Chia-Nan University of Pharmacy & Science, Tainan, Taiwan, ROC

Abstract

This study mines customer behavior to assist managers in developing better promotion and other relevant policies for a firm. The association rules of the relational database design are implemented in the mining system which provides electronic catalog designs and promotional policies. It provides marketing managers with a useful tool to rapidly search for valuable information based on customer transaction information. Thus, it enables marketing managers to rapidly establish marketing strategies to enhance sales and profit. It results in better sales for the firm a retail store mall.

Keywords: Data mining; Relational database; Marketing; Electronic catalog design

1. Introduction

Information Technology has provided many useful tools for customers and firms in handling their decision problems for the 21st century. For example, customers can easily use the Internet to access information on various products from numerous vendors. The information then influences their purchase decisions. Furthermore, firms can collect and analyze the customers' information in order to make better decisions in marketing policy through many types of information technology.

Marketing management can solidify ephemeral relationships with customers into long-term and fruitful relationships if it can discover and predict changes in customer behavior. In the past, researchers have generally applied statistical surveys to research customer behavior. Recently, data mining technology has been widely adopted to predict customer behavior (Giudici & Passerone, 2002; Song, Kim, & Kim, 2001).

Data mining is one of the most popular technologies that can discover potential customer knowledge from business databases in order to help firm make better decisions, which can automatically or semi-automatically mine and analyze large amounts of data in a business database (e.g., a transactions database), and is capable of detecting potential significant patterns and rules underlying the data. These patterns and rules are useful in sketching the behavior of consumer purchases or other consumer habits.

Data mining approaches include association rule mining, clustering, classification, estimation, etc. Association rule mining is widely used for analyzing the product items purchased by consumers. It is also used to support sales promotion and marketing segmentation (Changchien & Lu, 2001; Changchien, Lee, & Hsu, 2004). The association rule is represented by $X \rightarrow Y$ where X and Y are a set of items. This rule means that the transaction records in a business database that contain X tend to contain Y. A large number of valid algorithms for mining association rules have been proposed (for example, Agrawal, Imielinski, & Swami, 1993; Agrawal & Srikant, 1995; Anand, Patrick, Hughes, & Bell, 1998; Han & Fu, 1999; Mastsuzawa & Fukuda, 2000, etc.).

In this study, a mining system to detect customer behavior is proposed. The association rules from relational database design are utilized to mine consumer behavior in order to generate cross-selling proposals for an electronic catalog

* Corresponding author. Tel.: +886 6 2757575x53137; fax: +886 6 2759451.
E-mail address: linn@mail.ncku.edu.tw (C. Lin).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.08.028

design and marketing for a retailing stationery mall in Taiwan. The other sections of the study are arranged as follows. In Section 2, we describe the background of the case firm and summarize the goal of the research project of the case firm. Section 3 introduces the proposed data mining system, which contains a system framework and relational database design. Section 4 introduces the data mining process, result analysis, and electronic catalog design. Discussions and conclusions are presented in Section 5.

2. Profile of the case

The stationery mall was established in 1996. The mall specializes in writing and stationery products. It has been a leading provider of personalized stationery and related products for consumers. In the past decade, the firm has been proud of its products' quality and the value expected by customers. The mall's main customers are teachers and students.

From our interviews with the company's marketing managers, the catalog design is the fundamental promotional tool for the firm. The product catalog is designed by the marketing office based on decisions from department meetings. No input is gathered from the branch stores' directors. According to the promotion schedule, approximately one month is required for the marketing office to design and produce the catalog. Furthermore, the catalog is redesigned seasonally. For example, more page space is dedicated to pen and paper products at the beginning of the school semester due to the high demand for these products during this period of the year. Hence, at the beginning of the semester the catalog would contain a hundred kinds of pen and 30 types of paper, while it would only contain 50 kinds of pen and 10 types of paper during the latter periods of the semester. Upon the completion of the catalog, copies of the catalog are delivered to each branch two days before it is released to the public.

The catalog is designed and produced by the marketing office based on aggregated information from all branches, but branch directors are without authority to make decisions on catalog content. Hence the catalog contents cannot provide a customer-oriented segmentation service for the different branches' needs. Furthermore, the cross-selling promotion is still neglected in current promotion policy. In addition to the traditional paper catalog, they have another promotion mode: making advertisements for magazines and newspapers for the Branch Anniversary.

3. System

3.1. Framework

The framework of the system is shown in Fig. 1, in which the relational database management system is used to conduct the data mining which consists of four steps:

Step 1: Input the system code by certification and open the databases, which consist of the customer database, the stationery mall database, and the transaction database.
Step 2: Analyze the data according to the database.
Step 3: Design the electronic catalog according to the analysis result.
Step 4: E-mail the electronic catalog to the relevant branches and to the customers.

The data was collected from September 2004 to December 2004. The database of the system consists of three major parts: customer data, product data, and transaction data.

3.2. Relational database

The concept of the relational database was first proposed by Codd (1970) and represents interrelated data in the form of tables. The interrelated tables are used to represent relationships among the data, these relationships forming the main characteristics of the relational database. Data is organized as a collection of tables in which related data is represented by one or more common values in related tables. As long as the two tables share a common data element, the tables then yield the associated data stored in one table with data stored in another table. The tables appear similar to flat files, but the data in more than one file can be easily extracted and combined with Structured Query Language (SQL), which is the standard data manipulation language of the relational database management system (David & Wenny, 2002; Laudon & Laudon, 2003). Relational databases are ubiquitous in the information systems of modern firms'. By embedding reliably portrayed consumer patterns within a database of relational patterns, firms can build database management systems that support a variety of operational activities (Tsechansky, Pilskin, Rabinowitz, & Porath, 1999).

Many studies have shown that association rules in a relational database can provide a powerful method for mining knowledge in different applications (for example, Berzal, Cubero, Marı'n, & Serrano, 2001; Breault, Goodall, & Fos, 2002; Dabbas & Chen, 2001; Tseng, 2005). In this paper, relational database management system follows the systems development approach. In order to integrate the implementation of the system, we design and construct a relational database. The relational database is divided into three data functions and contains seven data tables based on database functions (see Table 1).

3.2.1. Entity–relationship model

There are three main database design models in relational database design: the conceptual database, the logical database, and the physical database. The conceptual database is the first step in designing a relational database; Fig. 2 is the entity and relationship (E–R) diagram of the conceptual database which provides a conceptual database

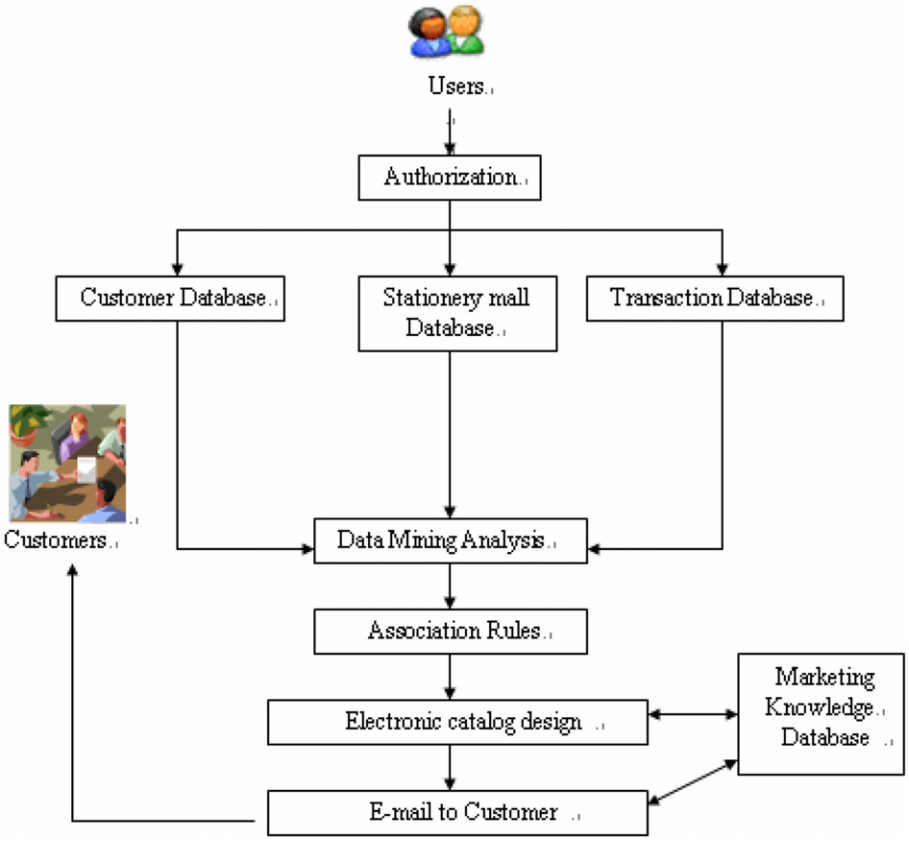

Fig. 1. System structure.

Table 1
Database functions and database tables

Database functions	Database tables
Customer data	Customer data Purchasing likes Purchasing data
Stationery mall data	Department data Area data of the department
Transaction data	Production data Production brand data

model for the case firm's relational database design. It is important to check the quality of such a design with the target users of the relational database application. We examined whether the E–R model could easily satisfy the user's demand for data or information. At this step, we discussed the E–R model and its functions with the marketing manager, financial manager, and chief of the MIS department so that we could determine whether their requirements would be satisfied by the information provided by E–R model. Finally, the E–R model is designed to include 7 entities, 9 relationships, and 42 attributes, as shown in Fig. 2.

3.2.2. Logical database design

The logical database design includes the logical database model, entities, attributes, and data relationships. Every entity is an object, event (e.g., a process), or concept. Data attributes are used to represent an entity's characteristics and the data relationship between entities, which explains the relationships between different objects and events inside the enterprise. The logical database model emphasizes 'logic', which is a readable method and useful for representing the knowledge. The logical database design is the data model that is most frequently used because it establishes a simple data form for each relation and for many-to-many relationships.

Logical database design is the process of transforming the conceptual database model into a logical database model that can be implemented on a chosen database management system. In general, the relational database model is used widely in designing modern database applications, and the principles of logical database design of a relational model can be easily applied to other logical models (Liu & Song, 2000; McFadden, Hoffer, & Prescott, 1999). Fig. 3 is a form of the logical database design that provides the data model of a relational database system in our case.

Moreover, the database functions and data tables of the case firm are described below:

(1) Customer data: consists of a customer information table, a customer purchasing preferences table, and a customer purchasing data table.
(2) Product data: the product database consists of a basic product information table, a product price table, and a brand table.

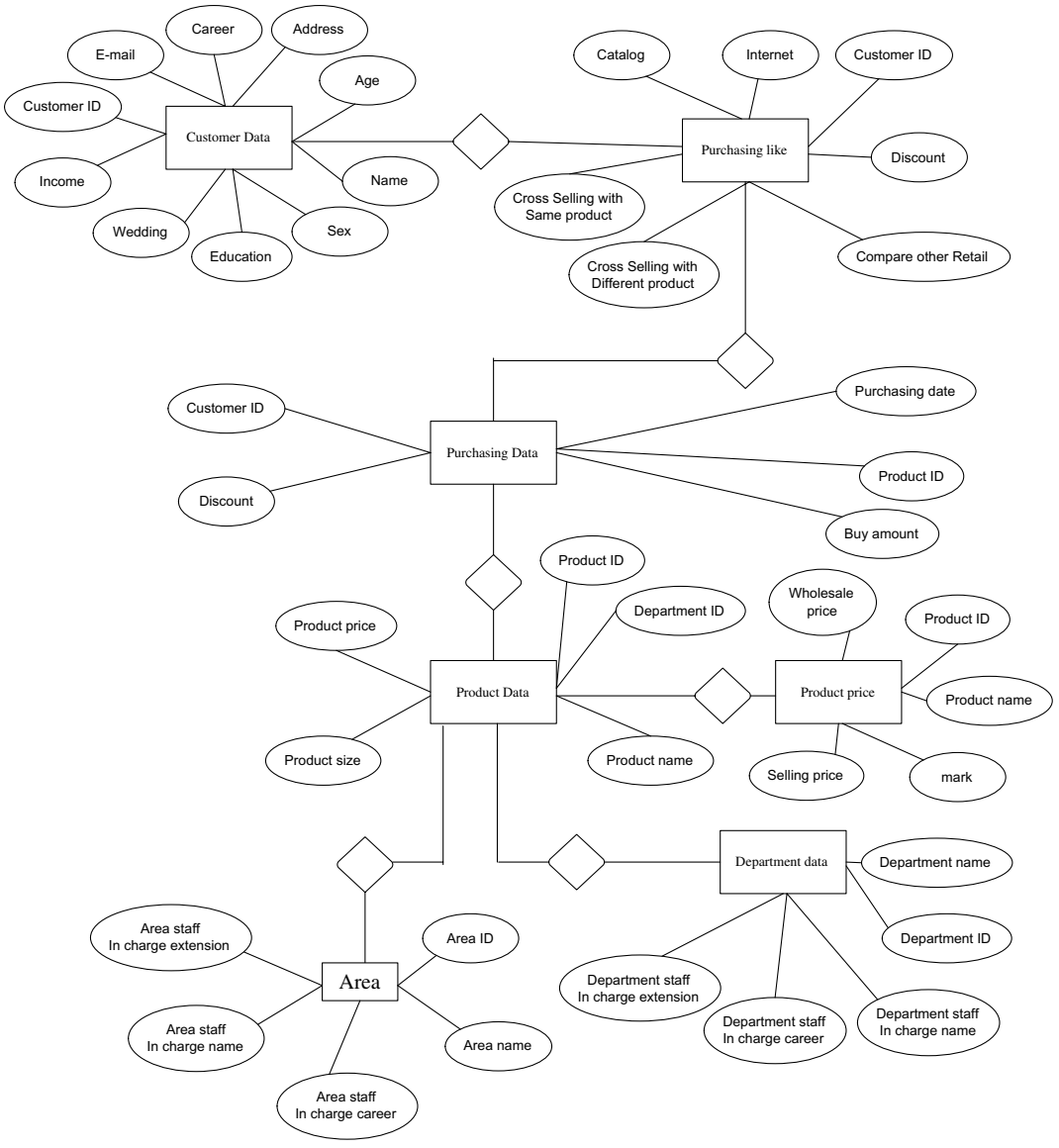

Fig. 2. E–R model.

(3) Stationery mall data: the stationery mall data includes a departmental list and a sectional list. With these lists the employee can easily and quickly revise the data of the selling display according to the consumers' needs. In addition, it can conveniently provide the required information to top management, who can quickly understand the situational condition of the sales market, and then make the right decision.

3.2.3. Physical database design

The main purpose of the physical database design is to transform the logical description into a technical description for storing and retrieving data. Thus, the physical database should incorporate a data storage design that will provide adequate performance and ensure database integrity, database security, and database recoverability. Once the physical database design is finished, the operating systems, database management systems, and the data access languages can obtain the best decision for implementation. In this paper, the structure of physical database design is described in Fig. 4.

4. Data mining system

4.1. Data mining procedures

Data mining procedures generally consist of database establishment, data mining function, segmentation analysis, knowledge acquisition, catalog marketing, and sales promotion. The data mining process in this research is divided into the following steps (as shown in Fig. 5).

The data mining process in the study includes the following steps:

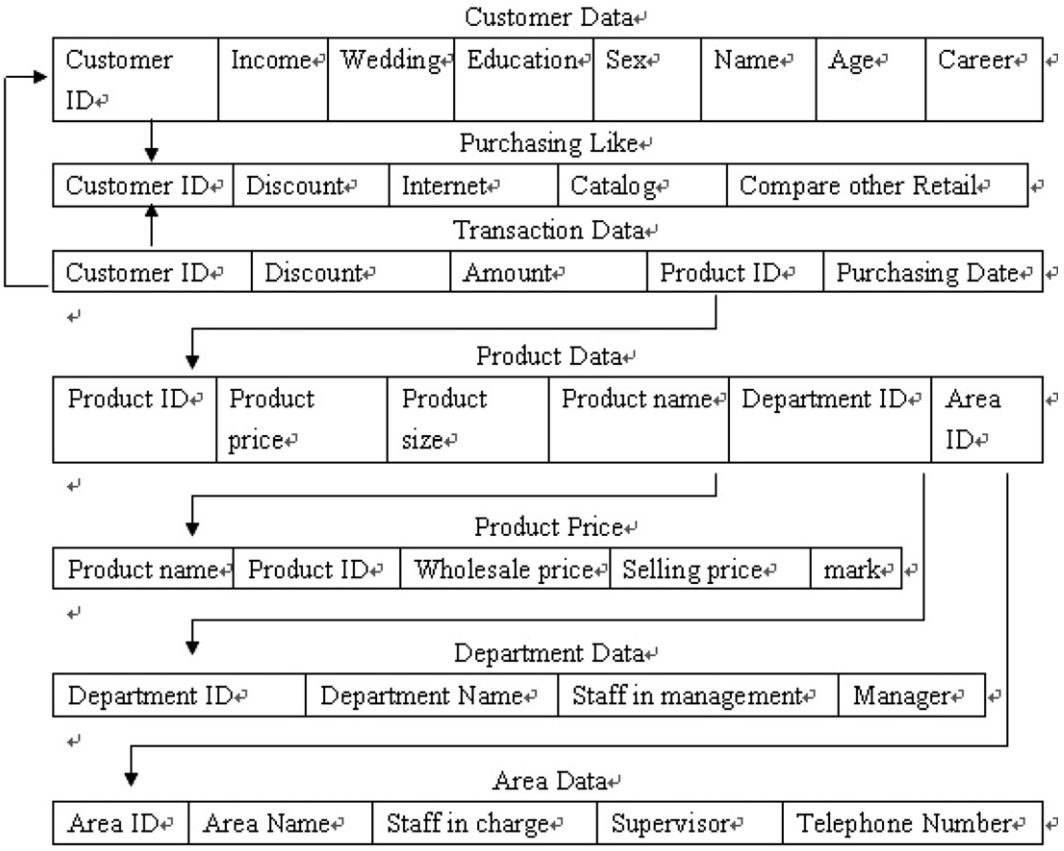

Fig. 3. Logical database design.

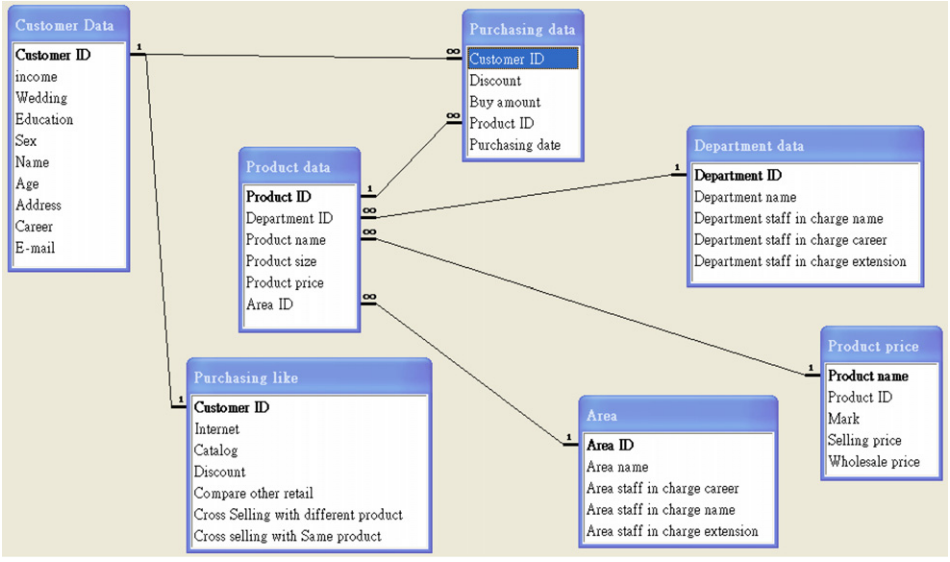

Fig. 4. Physical database design.

Step 1: database establishment.

We collected data from 09/01/2004 to 12/31/2004. The result was that 6583 customers' data were collected, of which 5876 are deemed valid. The career distribution is as follows: students 40%, teachers 26%, manufacturing industry 6%, service industry 8%, public officials 12%, information industry 5% and self-employed 3%, respectively. Because the case mall had not previously constructed a database, data were collected and input into data tables from fieldwork in order to build the relational

database, including a case mall database, customer database, and transaction database. To search for relationships between data tables from different data functions, the relational data in data functions could be associated from different data tables for data mining.

Step 2: data mining.

The data mining step implements association queries according to decision variables in order to analyze consumer behaviors so that the marketing/sales department can plan strategies and tactics for target consumers. The objective of database mining includes mining customer transaction data, mining purchased product collection, and mining brand collection of purchased products.

Step 3: segmentation analysis.

After database mining was implemented, customer purchasing patterns and their segmentation of product and brand were obtained from the database. By doing so, mining results from different relational tables can generate business information and knowledge.

Step 4: knowledge acquisition.

Some specific knowledge patterns or rules from mining results are acquired for building and maintaining a knowledge base. These knowledge patterns or rules could be a knowledge base in terms of customer relationship and knowledge management for marketing and sales management.

Step 5: electronic catalog design and sales promotion.

With knowledge of their customer, the case mall therefore started to design catalog content, including generalized format (paper catalog) and customized format (electronic catalog). By doing so, collections of target customers, target products, target brands, and target prices (discount) were designed and mailed to all customers or specific segmentation groups in order to implement catalog marketing. In addition, sales promotion of the case mall was also built according to mining knowledge so that marketing and sales management could be integrated.

Note that all product names appearing in the following section have been altered to maintain their confidentiality for the case firm.

4.2. Case examples

The first level of data mining uses the product department as decision variables, For example, we let *aa*, *ab*, *ac*, *ad*, denote the Pen department, the Notebook department, the Stapler department and the office 3C department, respectively. The equation $P(aa \cap ab = \text{Type1}|T)$ denotes the probability that the customers will buy products from the Pen department and Notebook department; where Type1 and T denote the product combination Type1 and

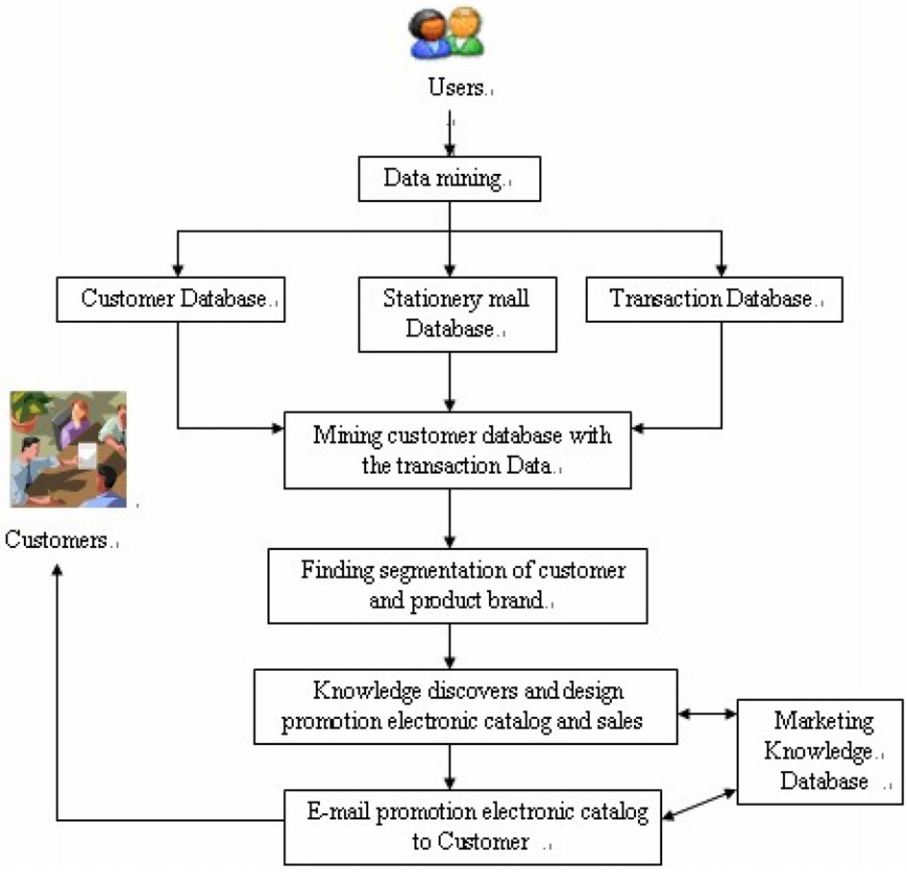

Fig. 5. Data mining procedure.

Table 2
Results of section classification as the decision variable (first level)

Product classification as decision variable (step 1)		
Product combination code	Product combination (%)	Rank
$aa \cap ab$	32.8	1
$aa \cap ac$	25.3	2
$ab \cap ae$	17.6	3
$ab \cap ac$	13.2	4
$ac \cap ad$	8.6	5

Table 3
Results of product brand (second level)

Product brand as decision variable (step 2)		
Product combination code	Product combination (%)	Rank
$aa2 \cap ab3$	62.8	1
$aa1 \cap ac2$	58.7	2
$ab1 \cap ae7$	56.3	3
$ab2 \cap ac2$	50.2	4
$ac5 \cap ad8$	38.6	5

customer, respectively. The detailed results are shown in Table 2.

The decision variable in the second level data mining is product brand. Herein, three product brands are sold by the notebook department (i.e., MONO notebook coded by *ab*3, COLOR notebook coded by *ab*1, and OFIS notebook coded by *ab*2), one product brand is sold by the Stapler department (i.e., STRONG stapling machine coded by *ac*2) and two product brands are sold by the Pen department (i.e., PILOT pen coded by aa2, and MITSUBISHI pen coded by *aa*1). The equation $P(aa2 \cap ab3 = \text{Type1})$ denotes the probability that the customers will buy *aa*2 and *ab*3 simultaneously. Please see Table 3 for more detailed results.

The decision variables in the third level of data mining are product brands with the product name. For example, the combination of the PILOT pen with the MONO color notebook ranks No. 1 in sales. Please see Tables 4 and 5 for detailed results.

4.3. Designing the electronic catalog from mined knowledge

According to the mining results, the promotion projects are to design some electronic catalogs for the customer in accordance with the customer's preference for the product brand to enhance product sales.

4.3.1. Promotion project A (i.e. include all brands)

Herein, the electronic promotional catalogs (i.e. Figs. 6 and 7) are designed following Rule 1 and Rule 2.

Rule 1: combine the sale, offering a 10% discount to all customers, of 3M notepaper with MONO color notebooks and PILOT pens, which rank No. 1 in product sales of Tables 4 and 5 respectively. Fig. 6 shows the description in the electronic catalog.

Rule 2: combine the sale of 3M notepaper with STRONG small staplers and MITSUBISHI pens, which rank No. 2 in product sales of Tables 4 and 5 respectively, offering a 15% discount to all customers. Fig. 7 shows the description in the electronic catalog.

The results of mining enabled the sales ranking of product brands as main promotional products and the designing of an electronic promotional catalog based on three products combined. Because some customers had bought two products at the same time, in order to encourage customers to purchase a third one, the third one, which belongs to the customers' product brands preference list, is offered in the catalog with a discount policy.

4.3.2. Promotion project B (i.e. only own brand)

Herein, the electronic promotional catalog is designed based on Rule 3.

Rule 3: combine the sale of 3M notepaper with the NANKING paper trimmer, offering a 10% discount to all customers. Fig. 8 shows the description in the electronic catalog.

The results of mining shows that selling own brand stationery products with best-selling 3M notepaper can raise sales volume of own brand products effectively and provide

Table 4
Results of product brand combine (third level)

Decision variable	Combined name of products	Rank
Stationery section's pen and notebook	a PILOT ball point pen paired with a MONO color notebook (62.8%)	1
Stationery section's pen and stapling machine	a MITSUBISHI ball point pen paired with a STRONG small stapling machine (58.7%)	2
Stationery section's pen, notebook, correction fluid, and paper cutter	PENTEL correction fluid, OFIS notebook, RABBIT ball point pen, and NANKING paper cutter are purchased by many consumers	

Table 5
Results of product brand combine (third level)

Decision variable	Combined name of products	Rank
Stationery section's office 3C and notebook	3M notepaper paired with MONO color notebook (42.36%)	1
Stationery section's office 3C and stapling machine	3M notepaper paired with a STRONG stapling machine (40.28%)	2
Stationery section's office 3C and pen	3M notepaper paired with RABBIT ball point pen (38.16%)	3

Fig. 6. Rule 1 electronic catalog design.

Fig. 7. Rule 2 electronic catalog design.

Fig. 8. Rule 3 of electronic catalog design.

another product combination for customers. To promote own brand products, the NANKING paper trimmer is selected as a main promotional product because customers have lower loyalty to it. In this way, it may greatly increase the sales volume of the NANKING paper trimmer and the loyalty to own brand products.

5. Discussion

This research provides a useful information system to enhance the sales of the case firm's products by changing the original promotion method. It helps supervisors and employees make better decisions about promotional activities by providing them with useful knowledge from the new system proposed here.

From the interview with the employees of the company, we found that (1) for the line employees, half of them were angry and frustrated with having to learn computer skills at the beginning of the training project for the new system; however, at the end of the training, they were satisfied due to a feeling of self-fulfillment; (2) for the supervisors, they were comfortable and tended to commit to the new system because it provided them with useful information that would facilitate the monitoring of routine activities done by line employees. It also helped the supervisors in making decisions concerning promotional action.

A review of the sales for the last three months, compared with the same months of the last year before implementing the system, showed that there was an increase in sales (see Fig. 9).

After applying electronic promotional catalogs in marketing, it is helpful to the company to catch on to the market acceptance of products and to know product sales conditions through the data mining system so as to mend marketing strategy and achieve the goal of product sales. Besides, the data mining system also can correct the design

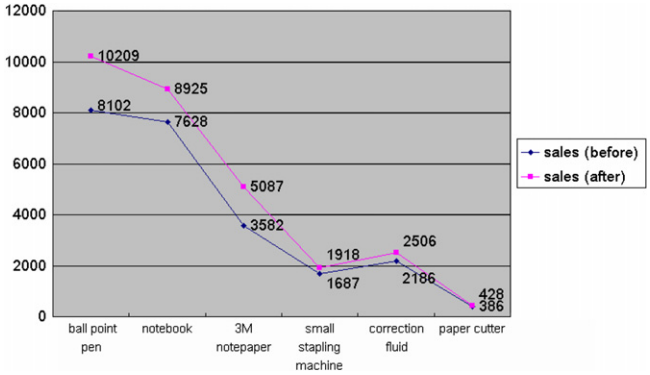

Fig. 9. Comparing sales before and after using new system.

of electronic catalogs according to market sales data and design electronic promotional catalogs of different periods. Thus, applying the data mining system effectively cannot only allow understanding of customers' purchasing behavior, but also provide knowledge of the cognition related to customers' purchasing for the policymaker to make the best policy.

This system provides a function that combines different but relevant products together in a promotion and then emails the electronic discount ticket to potential customers. This is in contrast to the system proposed by Changchien et al. (2004), which uses on-line customer data rather than including the customer data of the whole entity passageway. Its advantage resides in this on-line property while neglecting the whole structure of the customer data (Changchien et al., 2004).

Furthermore, this system provides feasible space for future development. For example, more detailed operation steps are left for the new user, such as integrating additional decision algorithms (e.g., neural networks, fuzzy theory, etc.), or using on-line analysis processing (OLAP), etc.

6. Conclusion

In this paper, we developed an information system that can detect changes in customer behavior based on customer transaction data and sales data. The advent of data mining has enhanced the accuracy of predicting customer behavior. Mining for changes in customer behavior is useful for satisfying customer needs in a dynamic business environment. In this study, the association rules are used to identify links between customer profiles and products purchased. It provides marketing managers with a useful tool to rapidly search for valuable information based on customer transaction information. Thus, it enables marketing managers to rapidly establish marketing strategies to enhance sales and profit.

Although we have proposed the promotion decision support system, the pricing strategy and the methods for clustering customers can be further enhanced in future studies. For example, pricing policy can be considered with that of other competitors' prices in determining the promo-

tional prices. And other decision analysis such as game theory can be employed to generate more effective and competitive prices for electronic commerce. As for customer clustering, clustering based directly on customers' profiles may not lead to good clustering results. Due to the diversity in individual consumer behavior, cognitive needs, and personality, further research on methods to cluster customers may be quite interesting and helpful. Lastly, since customers change over time, the use of dynamic data mining methods can efficiently analyze and adjust consumer behavior.

Acknowledgement

Thanks to the case firm who provided the useful data and permitted us to publish the key part and results of the project.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association between sets of items in massive database. In *International proceedings of the ACM-SIGMOD international conference on management of data* (pp. 207–216).

Agrawal, R., & Srikant, R. (1995). Fast algorithms for mining association rules. In *Proceedings of the international conference on very large data bases* (pp. 407–419).

Anand, S. S., Patrick, A. R., Hughes, J. G., & Bell, D. A. (1998). A data mining methodology for cross-sales. *Knowledge-Based Systems, 10*, 449–461.

Berzal, F., Cubero, J., Marı'n, N., & Serrano, J. (2001). TBAR: an efficient method for association rule mining in relational databases. *Data and Knowledge Engineering, 37*, 47–64.

Breault, J. L., Goodall, C. R., & Fos, P. J. (2002). Data mining a diabetic data warehouse. *Artificial Intelligence in Medicine, 26*, 37–54.

Changchien, S. W., Lee, C.-F., & Hsu, Y.-J. (2004). On-line personalized sales promotion in electronic commerce. *Expert Systems with Applications, 27*(1), 35–52.

Changchien, S. W., & Lu, T. C. (2001). Mining association rules procedures to support on-line recommendation by customers and products fragmentation. *Expert Systems with Applications, 20*(4), 325–335.

Codd, E. F. (1970). A relational model of data for large relational database. *Communications of the ACM, 13*, 377–387.

Dabbas, R. M., & Chen, H. N. (2001). Mining semiconductor manufacturing data for productivity improvement—an integrated relational database approach. *Computers in Industry, 45*, 29–44.

David, T., & Wenny, R. J. (2002). Parallel database sorting. *Journal of Information Sciences, 146*, 171–219.

Giudici, P., & Passerone, G. (2002). Data mining of association structures to model consumer behavior. *Computational Statistics and Data Analysis, 38*, 533–541.

Han, J., & Fu, Y. (1999). Mining multiple-level association rules in large databases. *IEEE Transactions on knowledge and data engineering, 11*(5), 798–804.

Laudon, K. C., & Laudon, J. P. (2003). *Essentials of management information systems* (5th ed.). New York: Prentice-Hall.

Liu, W. Y., & Song, N. (2000). A logical design method for temporal databases based on the stability constraint and the generalization semantics. *Journal of Information Sciences, 124*, 153–171.

Mastsuzawa, H., & Fukuda, T. (2000). Mining structured association patterns from databases. In *Proceeding of the 4th Pacific-Asia conference, PAKDD 2000* (pp. 233–244).

McFadden, F. R., Hoffer, J. A., & Prescott, M. B. (1999). *Modern database management* (5th ed.). New York: Addison-Wesley.

Song, H. S., Kim, J. K., & Kim, S. H. (2001). Mining the change of customer behavior in an Internet shopping mall. *Expert Systems with Applications, 21*, 157–168.

Tsechansky, M. S., Pilskin, N., Rabinowitz, G., & Porath, A. (1999). Mining relational patterns from multiple relational tables. *Decision Support Systems, 27*, 177–195.

Tseng, F. S. C. (2005). Design of a multi-dimensional query expression for document warehouses. *Journal of Information Sciences, 174*, 55–79.